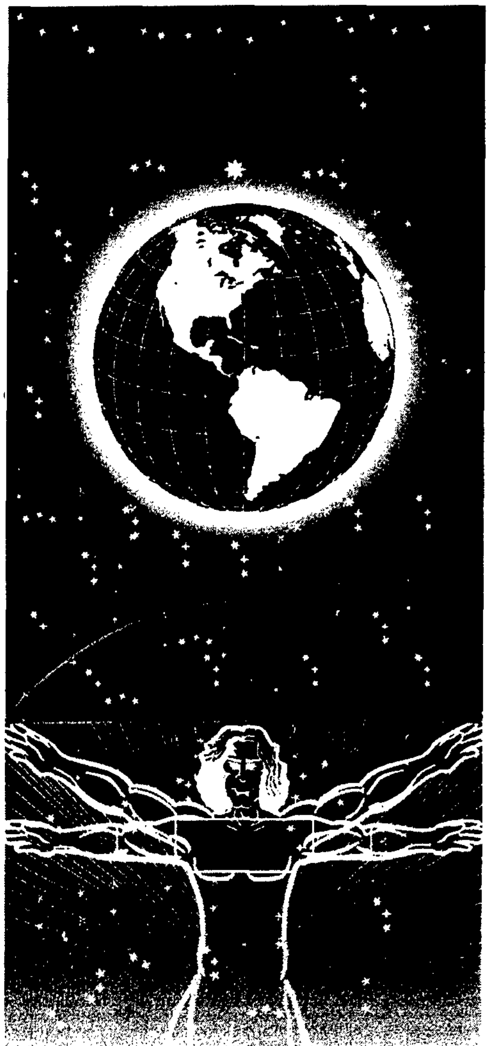
SAM-FE-BR-TR-2004-0001



UNITED STATES AIR FORCE
SCHOOL OF AEROSPACE MEDICINE

# Neuropsychological Evaluation of Aviators: Need for Aviator-Specific Norms?


William T. Thompson

Terra Health
Cypress Tower
1222 North Main Avenue
San Antonio, TX 78212

Daniel R. Orme, Major, USAF, BSC
Thomas M. Zazeckis, Lt Col, USAF, BSC




November 2004

Approved for public release;

distribution unlimited.


USAF School of Aerospace Medicine
Clinical Sciences Division
2507 Kennedy Circle
Brooks City-Base TX 78235-5116

## NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The mention of trade names or commercial products in this publication is for illustration purposes and does not constitute endorsement or recommendation for use by the United States Air Force.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

Government agencies and their contractors registered with Defense Technical Information Center (DTIC) should direct requests for copies to: Defense Technical Information Center, 8725 John J. Kingman Rd., STE 0944, Ft. Belvoir, VA 22060-6218.

Non-Government agencies may purchase copies of this report from: National Technical Information Services (NTIS), 5285 Port Royal Road, Springfield, VA 22161-2103.

Thomas M. Zazeckis

THOMAS M. ZAZECKIS, LtCol, USAF BSC
Chief, Medical Flight Screening Neuropsychiatry

DIANE C. RITTER, Col, USAF, MC, SFS
Chief, Clinical Science Division

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



**1. AGENCY USE ONLY** *(Leave blank)*

**2. REPORT DATE**
November 2004

**3. REPORT TYPE AND DATES COVERED**
Final

**4. TITLE AND SUBTITLE**
Neuropsychological Evaluation of Aviators: Need for Aviation-Specific Norms?

**5. FUNDING NUMBERS**

**6. AUTHOR(S)**
Orme, Daniel R.
Zazeckis, Thomas M.
*Thompson, William T.

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
*Terra Health
Cypress Tower
1222 North Main Avenue
San Antonio, TX 78211

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
USAF School of Aerospace Medicine
Clinical Sciences Division
2507 Kennedy Circle
Brooks City-Base, TX 78235-5116

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
SAM-FE-BR-TR-2004-0001

**11. SUPPLEMENTARY NOTES**

**12a. DISTRIBUTION AVAILABILITY STATEMENT**

Approved for public release; distribution unlimited.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** *(Maximum 200 words)*

To determine medical qualification to fly, aviators who sustain head trauma or acquire other conditions that effect mental skills must undergo neuropsychological evaluations. A challenge for psychologists tasked with performing these evaluations is that conventional neuropsychological tests are characteristically normed using samples reflecting the general population. It would be more appropriate to compare aviators with a sample of their peers. The present paper demonstrates the importance and usefulness of aviator-specific psychometric norms. To do this, normative tables were developed from a large sample of United States Air Force pilot training candidates using the Multidimensional Aptitude Battery, a standardized intelligence test. Psychologists may find these tables useful when evaluating aviators.

**14. SUBJECT TERMS**
Aviators intelligence; Pilot applicants' characteristics; Air Force undergraduate pilots; Normative data for pilot applicants; Multidimensional Aptitude Battery; MAB; Head Trauma

**15. NUMBER OF PAGES**
24

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | UL |

Standard Form 298 (Rev. 2-89) (EG)
Prescribed by ANSI Std. 239.18
Designed using Perform Pro, WHS/DIOR, Oct 94

**THIS PAGE INTENTIONALLY LEFT BLANK**

## TABLE OF CONTENTS

## LIST OF TABLES

## NEUROPSYCHOLOGICAL EVALUATION OF AVIATORS: NEED FOR AVIATOR-SPECIFIC NORMS?

## INTRODUCTION

Aviators who sustain head trauma or acquire illnesses that effect mental skills undergo neuropsychological evaluations in order to determine their medical qualification to return to flying. However, most standard neuropsychological tests are developed using normative samples reflecting the general population. Since it can be argued that aviators represent a unique population it is most appropriate that their performance on testing be compared with a sample of their peers. Few neuropsychological tests exist that use aviator norms. This presents a challenge for psychologists who are tasked with conducting these critical evaluations. The present paper discusses the need for aviator-specific norms and demonstrates their usefulness using intelligence test norms developed using a large sample of United States Air Force pilot training candidates.

The need for population-specific norms for psychological tests is well established. Grant and Adams (1996, p.142) noted "the purpose of normative data is to provide information on the range of an ability within a specifically defined population," adding that they should "be an unbiased sample of the population of interest." This is important since research repeatedly has shown relationships between demographic variables (e.g., age, education, gender) and psychological test results (Heaton, Grant, & Adams, 1991; Heaton, Ryan, Grant, & Matthews, 1996).

Since military aviators generally perform in excess of one standard deviation above the mean on standardized intelligence tests (Retzlaff, Callister, & King, 1999; Retzlaff & Gibertini, 1998) they should be considered, from a psychometric standpoint, to be a unique population. Consequently, when an aviator is given psychological testing it is most appropriate to evaluate these results in relation to those obtained by other aviators. However, little normative data, using tests commonly used in clinical settings, has been gathered on this population. Kay's (1995) computer-administered neuropsychological test battery is the main exception. This was developed for use with aviators and has norms based on a sample of commercial aviators. Still, this test is little known or used outside the aviation psychology community.

### Evaluation Process

Clinical neuropsychologists assess "brain function inferred from an individual's cognitive, sensory/motor, emotional, or social behaviors" (Howieson & Lezak, 1997, p. 181). They offer opinions concerning the presence of brain injury, localization of impairment, injury severity, neuropsychological strengths and weaknesses, ability to return to pre-injury activities, and to assist in rehabilitation planning (Franzen & Lovell, 1987). In order to do so they may use a wide variety of tools. These include detailed clinical interviews with patients and significant collateral contacts, behavioral observations, subjective tests (e.g., confrontation testing of visual fields), and psychometric testing to include personality, intellectual, and cognitive tests.

Interpretation of psychometric data involves two processes: use of normative data and pattern analysis. Through the use of normative data it is possible to view an individual's abilities compared with peers. With aviators who have sustained illness or injury that effects mental skills, should abilities be considerably lower than one's peers then there would be reason to suspect cognitive impairment and, possibly, decline. This is especially important if the performance is poor on tasks on which unimpaired individuals consistently do well such as naming common items. When skills are normally distributed in the population, cut-off scores are sometimes used to suggest when a performance is in the "impaired" range and this usually is when the score is greater than two standard deviations below the mean (Howieson & Lezak, 1997).

Pattern analysis, on the other hand, involves examining the patient's performance on a variety of tasks that require different skills. It assumes that "consistency in the expression of cognitive functions is a key concept" (Lezak, 1996, p. 167). Generally speaking, this means that non-brain-impaired individuals perform at a consistent level across a range of cognitive skills. If there is a significant difference in performance on tests that assess divergent skills then it could suggest that this represents a deficit due to brain impairment. Significant differences can be denoted either by statistically significant differences between scores or by the use of base rates. Base rates show the frequency of differences in scores obtained on two tests or subscales in the standardization sample. For example, a patient received Immediate Memory and Working Memory scores of 96 and 111, respectively, on the Wechsler Memory Scale – Third Edition (Wechsler, 1997). According to normative data, this 15 point difference is statistically significant at greater the $p<.05$, suggesting that the patient has better working memory than immediate memory. However, differences of this magnitude or greater were seen in 38% or the national normative sample. Consequently, it is not uncommon and the apparent skill differential is most likely not noteworthy from a clinical perspective. It is important, then, for the neuropsychologist to examine both statistical differences and base rate when doing pattern analysis. Also, it is important to note that differences between two scores does not indicate, with certainty, that there is brain injury. Rather, this suggests this is a possibility, especially if findings are consistent with known patterns of impairment associated with specific neurological conditions, and warrants further investigation.

Evidence of the importance of pattern analysis abound. For example, an individual whose oral communication skills are considerably poorer than other, non-expressive language skills may have an expressive aphasia. Similarly, consider an individual's performance on an intelligence test that consists of several subtests. That person may obtain a solid score on subtests that involve retrieval of well-learned information held in long-term storage (e.g., vocabulary and general information) but evidence difficulty when required to repeat from memory strings of digits of increasing length that are presented orally. In this case, it is unclear whether this variable performance is due to deficits of attention, repetition, or other causes. However, it is clear that further evaluation would be appropriate.

## METHOD

### Participants

All subjects (N = 5617) were either active-duty United States Air Force (USAF) personnel, Air Force Academy students, or recent Air Force Reserve Officer Training Corps (ROTC) graduates who had been selected to attend undergraduate pilot training. They had completed and passed Class I physical examinations, indicating they were in excellent physical health. Data was collected between April 1994 and August 1999 as part of the USAF Enhanced Flight Screening - Medical, a program designed to obtain baseline performance measures and test for aeromedically disqualifying conditions. All either were, or soon would be, college graduates. Mean age was 22.98 (SD = 2.44) and ranged from 19 to 34 years of age. Most were male (91.8%) and Caucasian (91.7% Caucasian; 2.7% Black; 2.4% Hispanic; 1.3% Asian; 1.9% Other). The voluntary, fully informed consent of the subjects used in this research was obtained as required by 32 CFR 219 and AFI 40-402.

### Materials

The Multidimensional Aptitude Battery (MAB) (Jackson, 1984) is a group-administered test of intelligence. It consists of ten subtests, each seven minutes long, and produces Verbal (VIQ), Performance (PIQ), and Full Scale (FSIQ) IQ scores as well as subscale scaled scores. This test is patterned after the Wechsler Adult Intelligence Scale – Revised (WAIS-R) (Wechsler, 1981); however, the WAIS-R is an individually-administered test. Additionally, and in contrast to the WAIS-R, all test items on the MAB are multiple-choice. While there are no test items in common between these tests, nine of the ten MAB subtests share names with those in the WAIS-R and, at face validity, appear to assess similar skills; one subtest (Spatial) has been substituted for Block Design. Correlations between WAIS-R subscales and their MAB counterparts range from .44 to .89 (Jackson, 1984) and are generally stronger than those associated with the WAIS-R subtests and their earlier versions on the original WAIS. Correlation between the WAIS-R and MAB FSIQ scores is .91; MAB FSIQ test-retest reliability is .97 (Jackson, 1984). The MAB assesses a variety of cognitive skills, including vocabulary, general knowledge, verbal and nonverbal abstract reasoning, and spatial analysis. It is commonly given and commercially available.

## RESULTS/DISCUSSION

As noted above, interpretation of psychometric data involves comparison of obtained scores with normative data and the use of pattern analysis. Normative data ideally should be based on a sample of the examinee's peers. This method, and its importance, will be demonstrated using MAB information obtained as a result of the present study.

Table 1 lists MAB VIQ, PIQ, FSIQ, and subscale mean and standard deviation scores by ethnic group and gender based on the sample used in this study. For aviators, comparison with a normative sample of peers is of particular importance when evaluating for possible sequelae of brain injury. For example, an evaluation that results in a FSIQ score of 105 is well within normal limits when compared with the MAB's national normative sample. However, this is approximately two standard deviations below the USAF pilot training candidate mean (Table 1). Given this, it may be suggestive of a decline in overall intellectual functioning and warrant a more comprehensive neuropsychological evaluation. It is not unusual, for example, to see a pilot's FSIQ score rise from 100 shortly after a head injury to the 120's range twelve months later. This increase suggests that the brain was injured sufficiently severe to result in a general cognitive decline but, with the passage of time, experienced spontaneous remission of symptoms.

**TABLE l. Mean and Standard Deviation VIQ, PIQ, FSIQ and Subscale Scores By Total Sample, Ethnic Group, and Gender**

| Variable | Total | Black | Caucasian | Hispanic | Asian | Male | Female |
|---|---|---|---|---|---|---|---|
| VIQ | 119.6(6.4) | 115.5(6.6) | 119.6(6.3) | 116.0(7.5) | 119.3(5.5) | 119.8(6.4) | 118.0(6.0) |
| PIQ | 118.5(8.2) | 112.6(8.9) | 118.8(8.9) | 116.3(8.5) | 118.7(7.7) | 118.7(8.2) | 116.3(8.3) |
| FSIQ | 120.8(8.2) | 115.1(6.8) | 120.7(6.3) | 117.3(7.3) | 120.4(5.5) | 120.7(6.4) | 118.4(6.2) |
| Inf | 66.9(6.0) | 64.0(6.5) | 67.1(6.0) | 64.7(7.2) | 67.4(5.6) | 67.1(6.0) | 65.1(5.8) |
| Com | 59.4(4.0) | 57.8(4.4) | 59.8(3.9) | 57.3(5.6) | 59.0(3.9) | 59.7(4.0) | 59.3(3.8) |
| Ari | 61.1(6.3) | 58.0(6.0) | 61.2(6.2) | 59.0(7.2) | 59.8(4.8) | 61.3(6.3) | 58.9(5.8) |
| Sim | 60.7(4.4) | 58.7(5.0) | 60.8(4.3) | 59.0(5.0) | 61.5(4.1) | 60.7(4.5) | 60.7(3.9) |
| Voc | 60.6(6.7) | 58.7(6.5) | 60.8(6.7) | 58.0(6.9) | 59.8(6.7) | 60.7(6.7) | 60.0(6.9) |
| DS | 63.8(6.6) | 60.7(7.4) | 63.9(6.6) | 62.7(6.8) | 64.8(6.1) | 63.6(6.7) | 65.5(6.3) |
| PC | 59.6(6.2) | 56.8(6.4) | 59.8(6.1) | 57.9(6.2) | 58.5(6.9) | 59.9(6.2) | 56.6(5.9) |
| SP | 59.9(6.7) | 57.4(6.6) | 60.0(6.6) | 59.4(6.9) | 59.6(6.4) | 60.1(6.6) | 57.5(7.2) |
| PA | 52.8(6.9) | 49.1(6.6) | 52.9(6.8) | 51.0(7.4) | 53.4(6.0) | 52.8(6.9) | 52.2(7.1) |
| OA | 59.8(5.6) | 56.0(6.9) | 60.0(5.5) | 59.2(5.3) | 60.5(5.0) | 59.9(5.5) | 58.6(5.9) |

FSIQ = Full Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; Inf = Information; Com = Comprehension; Ari = Arithmetic; Sim = Similarities; Voc = Vocabulary; DS = Digit Symbol; PC = Picture Completion; SP = Spatial; PA = Picture Arrangement; OA = Object Assembly

It is also important to note that the various ethnic groups did not consistently score similarly on all scales. SAS GLM MANOVA procedure was used to analyze the main effect of ethnic groups for FSIQ, VIQ, PIQ, and each subscale. Contrast estimates were performed to determine significant differences for each test, relative to ethnic group, where a significant main effect for ethnic group was determined. MAB IQ and subscale mean score differences were noted between the different ethnic groups and this is presented in Table 2. Thus, the Black and Hispanic groups performed similarly on the MAB as did the Caucasian and Asian groups. The "Other" group most closely resembled the Caucasian and Asian groups, suggesting it was largely made up of individuals from these groups.

**TABLE 2. Significant (p<.05) Mean Score Differences Between Ethnic Groups**

| | Black | Caucasian | Hispanic | Asian |
|---|---|---|---|---|
| Caucasian | FSIQ<br>VIQ<br>PIQ<br>Inf<br>Com<br>Ari<br>Sim<br>Voc<br>DS<br>PC<br>SP<br>PA<br>OA | | | |
| Hispanic | FSIQ<br>PIQ<br>OA | FSIQ<br>VIQ<br>Inf<br>Com<br>Ari<br>Sim<br>Voc | | |
| Asian | FSIQ<br>VIQ<br>VIQ<br>Inf<br>Sim<br>DS<br>PA<br>OA | ----- | FSIQ<br>VIQ<br>Sim | |
| Other | FSIQ<br>VIQ<br>PIQ<br>Inf<br>Ari<br>DS<br>PA<br>OA | ----- | FSIQ<br>VIQ<br>Inf<br>Com | ----- |

FSIQ = Full Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; Inf = Information; Com = Comprehension; Ari = Arithmetic; Sim = Similarities; Voc = Vocabulary; DS = Digit Symbol; PC = Picture Completion; SP = Spatial; PA = Picture Arrangement; OA = Object Assembly

That there were differences between the groups is not surprising since it is well known that average scores on psychological tests sometimes vary with different ethnic groups. Still, this difference highlights the need for ethnic specific norms. For example, the Black sample's mean FSIQ was 5.7 points lower than the Caucasian. Suppose, for example, that a Black pilot is being evaluated for possible decline in functioning as a result of head trauma and receives a FSIQ score of 108. This score is nearly two standard deviations below the overall pilot training candidate mean. Such a score may be suggestive, then, of a decline in functioning. However, when compared with the Black sample this score is at the 16th percentile (one standard deviation below the mean) which is in the low average range. Use of these tables, then, may provide more appropriate comparative standards than are currently available and reduce the risk of over-diagnosis of impairment in minority aviators.

Pattern analysis, on the other hand, involves examining an individual's performance across a range of tasks. These tasks can be different tests or subtests within one test. With the MAB, pattern analysis involves comparing performance on the different subtests. Inter-subtest score scatter analysis can reveal divergent cognitive strengths that can be suggestive of compromise of brain functioning. For example, knowledge of word meanings and speed of information processing are often thought to be resistant and vulnerable, respectively, to brain injury. On the MAB, these are assessed with the Vocabulary and Digit Symbol subtests. Evidence suggestive of compromise of cognitive functioning would exist should the Vocabulary score be considerably higher than Digit Symbol score. In this case, further evaluation would be appropriate.

However, when are differences large enough to be considered of clinical significance? In other words, how great a difference is required so that it cannot be attributed to chance? There are two general approaches to this: statistical significance and base rates.

Statistical significance means that the magnitude of the difference between two scores is of such magnitude that the probability of this being due to chance is minimal. Table 3 presents the MAB VIQ-PIQ difference scores that are required for statistical significance at the .05 and .01 levels by ethnic group and gender.

**TABLE 3. Magnitude of VIQ-PIQ Difference Required for Statistical Significance By Total Sample, Ethnic Group, and Gender**

| Significance | Black | Caucasian | Hispanic | Asian | Other | Female | Male |
|---|---|---|---|---|---|---|---|
| p < .05 | 3.77 | 3.48 | 3.85 | 3.20 | 4.02 | 3.47 | 3.53 |
| P < .01 | 4.92 | 4.55 | 5.02 | 4.18 | 5.25 | 4.53 | 4.62 |

For example, suppose a female pilot training candidate received a VIQ score of 125 and a PIQ of 115. The VIQ–PIQ difference score of 10 is significant at greater than the .01 level and suggests, generally speaking, stronger ability on tasks that require verbal mediation than those that are more heavily dependent upon visuomotor skills.

Similarly, significance of differences between subscale scaled scores can also be determined. Table 4 reveals, for example that an Information subtest score that is four points higher than that obtained on the Spatial subtest is significant at the .01 level. Note that Table 4 was calculated using the total sample. Additional tables were calculated for each ethnic and gender group (see Tables 6 through 12). The difference between scores required for significance was computed from the standard error of the difference (SEMdiff). Multiplying the standard error of measurement of the difference by an appropriate z value results in the amount of difference required for statistical significance at a given level of confidence. This is the same procedure that was used with the Wechsler Adult Intelligence Scale – Third Edition (Psychological Corporation, 1997) for computing significant differences.

**TABLE 4. Magnitude of Subscale Difference Required for Significance: All Groups**

| | Inf | | Com | | Ari | | Sim | | Voc | | DS | | PC | | SP | | PA | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | .05 | .01 | .05 | .01 | .05 | .01 | .05 | .01 | .05 | .01 | .05 | .01 | .05 | .01 | .05 | .01 | .05 | .01 |
| Com | 2.5 | 3.2 | | | | | | | | | | | | | | | | |
| Ari | 3.0 | 3.9 | 2.5 | 3.3 | | | | | | | | | | | | | | |
| Sim | 2.5 | 3.3 | 2.0 | 2.6 | 2.6 | 3.4 | | | | | | | | | | | | |
| Voc | 3.1 | 4.0 | 2.6 | 3.5 | 3.1 | 4.1 | 2.7 | 3.6 | | | | | | | | | | |
| DS | 3.0 | 4.0 | 2.6 | 3.4 | 3.1 | 4.0 | 2.7 | 3.5 | 3.2 | 4.2 | | | | | | | | |
| PC | 2.9 | 3.8 | 2.5 | 3.3 | 3.0 | 3.9 | 2.6 | 3.4 | 3.1 | 4.1 | 3.1 | 4.0 | | | | | | |
| SP | 3.1 | 4.0 | 2.6 | 3.4 | 3.1 | 4.1 | 2.7 | 3.6 | 3.2 | 4.2 | 3.2 | 4.2 | 3.1 | 4.0 | | | | |
| PA | 3.1 | 4.1 | 2.7 | 3.5 | 3.2 | 4.1 | 2.8 | 3.6 | 3.3 | 4.3 | 3.3 | 4.2 | 3.2 | 4.1 | 3.3 | 4.3 | | |
| OA | 2.8 | 3.6 | 2.3 | 3.0 | 2.8 | 3.7 | 2.4 | 3.2 | 3.0 | 3.9 | 2.9 | 3.8 | 2.8 | 3.7 | 2.9 | 3.8 | 3.0 | 3.9 |

FSIQ = Full Scale IQ; VIQ = Verbal IQ; PIQ = Performance IQ; Inf = Information; Com = Comprehension; Ari = Arithmetic; Sim = Similarities; Voc = Vocabulary; DS = Digit Symbol; PC = Picture Completion; SP = Spatial; PA = Picture Arrangement; OA = Object Assembly

Finally, score pattern analysis may be undertaken by examining difference scores in terms of not only whether they are statistically significant but also whether the magnitude of the difference score was seen commonly in this sample. This is analysis of the base rate (see Table 5). For example, a VIQ-PIQ difference score of six is significant at the .01 level for Hispanic subjects (Table 3) but base rate information reveals that a difference of this magnitude and greater was seen by 58.2% of the sample (Table 5). Consequently, this difference is not clinically meaningful. On the other hand, a VIQ-PIQ difference of 18 is statistically significant for Hispanic subjects (Table 3); also, a difference of this size or greater was seen in only 2.2% of this study's sample (Table 5). A difference of this magnitude, then, could be clinically meaningful and would warrant further evaluation of the aviator.

**TABLE 5. Base Rates for VIQ-PIQ Difference By Total Sample, Ethnic Group, and Gender**

| VIQ-PIQ | Percent of sample with as great or less VIQ-PIQ difference score | | | | | | |
|---|---|---|---|---|---|---|---|
| Difference | Total | Black | Caucasian | Hispanic | Asian | Female | Male |
| 0 | 5.0 | 4.7 | 5.0 | 2.2 | 7.2 | 6.2 | 4.8 |
| 1 | 14.6 | 16.9 | 14.6 | 9.7 | 20.3 | 15.7 | 14.5 |
| 2 | 24.2 | 23.6 | 24.4 | 23.1 | 27.5 | 24.6 | 24.2 |
| 3 | 33.3 | 30.4 | 33.1 | 35.8 | 42.0 | 31.9 | 33.4 |
| 4 | 41.7 | 37.8 | 41.7 | 42.5 | 44.9 | 41.5 | 41.7 |
| 5 | 49.4 | 43.2 | 49.5 | 54.5 | 49.3 | 47.7 | 49.6 |
| 6 | 57.6 | 50.7 | 57.9 | 58.2 | 59.4 | 56.5 | 57.7 |
| 7 | 64.0 | 56.1 | 64.3 | 66.4 | 62.3 | 63.2 | 64.1 |
| 8 | 70.2 | 60.1 | 70.6 | 70.9 | 71.0 | 71.0 | 70.1 |
| 9 | 75.1 | 64.9 | 75.5 | 73.9 | 75.4 | 74.9 | 75.1 |
| 10 | 79.1 | 68.2 | 79.5 | 79.9 | 78.3 | 78.7 | 79.2 |
| 11 | 82.9 | 72.3 | 83.2 | 83.6 | 82.6 | 82.3 | 82.9 |
| 12 | 86.3 | 80.4 | 86.4 | 85.8 | 84.1 | 85.1 | 86.4 |
| 13 | 89.4 | 87.8 | 89.4 | 88.8 | 89.9 | 89.1 | 89.4 |
| 14 | 91.4 | 89.2 | 91.4 | 91.0 | 91.3 | 90.7 | 91.5 |
| 15 | 93.4 | 90.5 | 93.5 | 91.8 | 92.8 | 82.7 | 93.5 |
| 16 | 94.7 | 93.9 | 94.7 | 94.8 | 95.7 | 94.2 | 94.8 |
| 17 | 95.8 | ---- | 95.9 | 95.5 | 97.1 | 95.8 | 95.8 |
| 18 | 97.8 | 95.9 | 97.0 | 97.8 | 98.6 | 97.1 | 97.0 |
| 19 | 97.6 | 96.6 | 97.6 | ---- | >99.0 | 97.6 | 97.6 |
| 20 | 98.1 | 97.3 | 98.1 | 98.5 | >99.9 | 98.0 | 98.1 |
| 21 | 98.5 | 98.0 | 98.5 | >99.0 | >99.0 | 98.2 | 98.6 |
| 22 | 99.0 | 98.6 | 99.0 | >99.0 | >99.0 | 98.9 | 99.0 |
| 23+ | >99.0 | >99.0 | >99.0 | >99.0 | >99.0 | >99.0 | >99.0 |

There are several limitations to this study. First, the data was collected on individuals who were selected for undergraduate pilot training. This may not be representative of pilots in general since not all candidates will successfully complete training. However, the vast majority of those selected for pilot training (around 85%) do graduate from training and, in fact, there is no widely accepted standard for failure from pilot training; this is often based on the subjective, although learned, opinion of the instructor pilot. Still, it could be that those who fail score differently on the MAB than those who are successful and including their scores in the normative data alters it somewhat. Secondly, the sample consisted of young adults. The tables, then, should be used with caution when assessing older individuals. Finally, the MAB was modeled after the WAIS-R. The WAIS-R has since been revised and includes a new normative sample (Psychological Corporation, 1997). Scores are slightly reduced on the WAIS-III when compared with the WAIS-R. A similar performance on both tests would result in VIQ, PIQ, and FSIQ scores that are 1.2, 4.8, and 2.9 points lower, respectively, on the WAIS-III than on the WAIS-R. Consequently, MAB scores may slightly overestimate ability based on current customary norms.

## CONCLUSIONS

As this study demonstrated, aviators perform considerably better on standardized psychometric testing than the general population. Determining whether an aviator has experienced a decline in cognitive skills following illness or injury, then, requires comparison of that individual's performance on testing with a sample of peers, in this case, other aviators. Use of normative data obtained from samples representing the United States population as a whole runs the risk of not identifying true decline in mental skills (i.e., false negative). Importantly, such a diagnostic "miss" could result in allowing an impaired aviator into the cockpit, placing that individual, other aircrew, and the mission in jeopardy. Since aviators represent a unique population, evaluations of their cognitive skills, particularly when a determination regarding returning to flying is to be made, should use instruments that have aviator-specific norms.

## REFERENCES


Franzen, M. D., & Lovell, M. R. (1987). Neuropsychological assessment. In S. Yudofsky & R. Hales (Eds.). The American Psychiatric Press textbook of Neuropsychiatry (3rd ed.). Washington (DC): American Psychiatric Press.

Grant, I., & Adams, K. (1996). Neuropsychological assessment of neuropsychiatric disorders. New York: Oxford University Press.

Heaton, R., Grant, I., & Adams, K. (1991). Comprehensive norms for an expanded Halstead-Reitan battery: demographic corrections, research findings, and clinical applications. Odessa, Florida: Psychological Assessment Resources, Inc.

Heaton, R., Ryan, L., Grant, I., & Matthews, C. (1996). Demographic influences on neuropsychological test performance. In: I. Grant & K. Adams (Eds.), Neuropsychological assessment of neuropsychiatric disorders (2nd ed.). New York: Oxford University Press.

Howieson, D., & Lezak, M. (1997). The neuropsychological evaluation. In S. Yudofsky & R. Hales, Neuropsychological assessment of neuropsychiatric disorders (2nd ed). New York: Oxford University Press.

Jackson, D. (1984). Multidimensional Aptitude Battery manual. Ontario, Canada: Research Psychologists Press.

Kay, G. (1995). CogScreen: Aeromedical Edition. Odessa, Florida: Psychological Assessment Resources, Inc.

Lezak, M. (1995). Neuropsychological assessment (3rd ed). New York: Oxford University Press. Psychological Corporation (1997). WAIS-III/WMS-III technical manual. San Antonio, Texas, Author.

Retzlaff, P., Callister, J., & King, R. (1999). Clinical procedures for the neuropsychological evaluation of U.S. Air Force pilots. Military Medicine, 164, 1-6.

Retzlaff, P., Gibertini, M. (1998). Objective psychological testing of U.S. Air Force officers in pilot training. Aviation, Space, and Environmental Medicine, 661-663.

Wechsler, D. (1981). WAIS-R manual. San Antonio, Texas: The Psychological Corporation.

Weschsler, D. (1997). The Wechsler Adult Intelligence Scale—Third Edition. San Antonio, TX: The Psychological Corporation.

# APPENDIX A – SIGNIFICANCE OF SUBSCALE DIFFERENCE SCORES BY RACE AND GENDER

**TABLE A-1. Magnitude of SubScale Difference Required for .05 and .01 Level of Significance: Blacks**

| | Inf | | Com | | Ari | | Sim | | Voc | | Ds | | Pc | | Sp | | Pa | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| alpha | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 |
| Com | 2.7 | 3.5 | | | | | | | | | | | | | | | | |
| Ari | 3.0 | 3.9 | 2.5 | 3.3 | | | | | | | | | | | | | | |
| Sim | 2.8 | 3.6 | 2.3 | 3.0 | 2.6 | 3.5 | | | | | | | | | | | | |
| Voc | 3.1 | 4.1 | 2.7 | 3.5 | 3.0 | 3.9 | 2.8 | 3.6 | | | | | | | | | | |
| Ds | 3.3 | 4.4 | 2.9 | 3.8 | 3.2 | 4.2 | 3.0 | 4.0 | 3.3 | 4.4 | | | | | | | | |
| Pc | 3.1 | 4.0 | 2.6 | 3.4 | 3.0 | 3.9 | 2.7 | 3.6 | 3.1 | 4.0 | 3.3 | 4.3 | | | | | | |
| Sp | 3.2 | 4.1 | 2.7 | 3.5 | 3.0 | 4.0 | 2.8 | 3.7 | 3.1 | 4.1 | 3.4 | 4.4 | 3.1 | 4.1 | | | | |
| Pa | 3.2 | 4.1 | 2.7 | 3.5 | 3.0 | 4.0 | 2.8 | 3.7 | 3.1 | 4.1 | 3.4 | 4.4 | 3.1 | 4.1 | 3.2 | 4.2 | | |
| Oa | 3.2 | 4.2 | 2.8 | 3.6 | 3.1 | 4.0 | 2.9 | 3.8 | 3.2 | 4.2 | 3.4 | 4.5 | 3.2 | 4.2 | 3.2 | 4.2 | 3.2 | 4.2 |

Inf=Information; Com=Comprehension; Ari=Arithmetic; Sim=Similerities; Voc=Vocabulary; DS=Digit Symbol; PC=Picture Comprehension; SP=Spatial PA=Picture Arrangement; OA=Object Assembly

**TABLE A-2. Magnitude of SubScale Difference Required for .05 and .01 Level of Significance: Caucasian**

| | Inf | | Com | | Ari | | Sim | | Voc | | Ds | | Pc | | Sp | | Pa | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| alpha | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 |
| Com | 2.4 | 3.2 | | | | | | | | | | | | | | | | |
| Ari | 2.9 | 3.8 | 2.5 | 3.2 | | | | | | | | | | | | | | |
| Sim | 2.5 | 3.3 | 2.0 | 2.6 | 2.6 | 3.4 | | | | | | | | | | | | |
| Voc | 3.0 | 4.0 | 2.6 | 3.4 | 3.1 | 4.0 | 2.7 | 3.5 | | | | | | | | | | |
| Ds | 3.0 | 3.9 | 2.6 | 3.4 | 3.1 | 4.0 | 2.7 | 3.5 | 3.2 | 4.2 | | | | | | | | |
| Pc | 2.9 | 3.8 | 2.5 | 3.2 | 3.0 | 3.9 | 2.6 | 3.3 | 3.1 | 4.0 | 3.1 | 4.0 | | | | | | |
| Sp | 3.0 | 4.0 | 2.6 | 3.4 | 3.1 | 4.0 | 2.7 | 3.5 | 3.2 | 4.2 | 3.2 | 4.2 | 3.1 | 4.0 | | | | |
| Pa | 3.1 | 4.0 | 2.7 | 3.5 | 3.1 | 4.1 | 2.8 | 3.6 | 3.2 | 4.2 | 3.2 | 4.2 | 3.1 | 4.1 | 3.2 | 4.2 | | |
| Oa | 2.8 | 3.6 | 2.3 | 3.0 | 2.8 | 3.7 | 2.4 | 3.1 | 2.9 | 3.8 | 2.9 | 3.8 | 2.8 | 3.7 | 2.9 | 3.8 | 3.0 | 3.9 |

Inf=Information; Com=Comprehension; Ari=Arithmetic; Sim=Similarities; Voc=Vocabulary; DS=Digit Symbol; PC=Picture Comprehension; SP=Spatial PA=Picture Arrangement; OA=Object Assembly

**TABLE A3. Magnitude of SubScale Difference Required for .05 and .01 Level of Significance: Hispanics**

| | Inf | | Com | | Ari | | Sim | | Voc | | Ds | | Pc | | Sp | | Pa | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| alpha | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 |
| Com | 3.1 | 4.0 | | | | | | | | | | | | | | | | |
| Ari | 3.4 | 4.5 | 3.1 | 4.0 | | | | | | | | | | | | | | |
| Sim | 3.0 | 3.9 | 2.6 | 3.3 | 3.0 | 3.9 | | | | | | | | | | | | |
| Voc | 3.4 | 4.4 | 3.0 | 4.0 | 3.4 | 4.4 | 2.9 | 3.8 | | | | | | | | | | |
| Ds | 3.4 | 4.4 | 3.0 | 3.9 | 3.4 | 4.4 | 2.9 | 3.7 | 3.3 | 4.3 | | | | | | | | |
| Pc | 3.2 | 4.2 | 2.8 | 3.7 | 3.2 | 4.2 | 2.7 | 3.5 | 3.2 | 4.1 | 3.1 | 4.1 | | | | | | |
| Sp | 3.4 | 4.4 | 3.0 | 4.0 | 3.4 | 4.4 | 2.9 | 3.8 | 3.3 | 4.4 | 3.3 | 4.3 | 3.2 | 4.1 | | | | |
| Pa | 3.5 | 4.6 | 3.2 | 4.1 | 3.5 | 4.6 | 3.0 | 4.0 | 3.5 | 4.5 | 3.4 | 4.5 | 3.3 | 4.3 | 3.5 | 4.5 | | |
| Oa | 3.0 | 4.0 | 2.6 | 3.4 | 3.0 | 4.0 | 2.5 | 3.2 | 3.0 | 3.9 | 2.9 | 3.8 | 2.8 | 3.6 | 3.0 | 4.1 | 3.1 | 4.1 |

Inf=Information; Com=Comprehension; Ari=Arithmetic; Sim=Similarities; Voc=Vocabulary; DS=Digit Symbol; PC=Picture Comprehension; SP=Spatial PA=Picture Arrangement; OA=Object Assembly

**TABLE A-4. Magnitude of SubScale Difference Required for .05 and .01 Level of Significance: Oriental**

| | Inf | | Com | | Ari | | Sim | | Voc | | Ds | | Pc | | Sp | | Pa | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| alpha | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 |
| Com | 2.3 | 3.0 | | | | | | | | | | | | | | | | |
| Ari | 2.5 | 3.3 | 2.1 | 2.7 | | | | | | | | | | | | | | |
| Sim | 2.3 | 3.1 | 1.9 | 2.5 | 2.1 | 2.8 | | | | | | | | | | | | |
| Voc | 3.0 | 3.9 | 2.6 | 3.5 | 2.8 | 3.7 | 2.7 | 3.5 | | | | | | | | | | |
| Ds | 2.8 | 3.7 | 2.4 | 3.2 | 2.6 | 3.4 | 2.5 | 3.2 | 3.1 | 4.0 | | | | | | | | |
| Pc | 3.0 | 3.9 | 2.7 | 3.5 | 2.9 | 3.7 | 2.7 | 3.6 | 3.3 | 4.3 | 3.1 | 4.1 | | | | | | |
| Sp | 2.9 | 3.8 | 2.5 | 3.3 | 2.7 | 3.6 | 2.6 | 3.4 | 3.2 | 4.1 | 3.0 | 3.9 | 3.2 | 4.2 | | | | |
| Pa | 2.8 | 3.6 | 2.4 | 3.2 | 2.6 | 3.4 | 2.5 | 3.2 | 3.1 | 4.0 | 2.9 | 3.8 | 3.1 | 4.0 | 3.0 | 3.9 | | |
| Oa | 2.5 | 3.3 | 2.1 | 2.8 | 2.3 | 3.1 | 2.2 | 2.9 | 2.8 | 3.7 | 2.7 | 3.5 | 2.9 | 3.8 | 2.8 | 3.6 | 2.6 | 3.4 |

Inf=Information; Com=Comprehension; Ari=Arithmetic; Sim=Similarities; Voc=Vocabulary; DS=Digit Symbol; PC=Picture Comprehension; SP=Spatial PA=Picture Arrangement; OA=Object Assembly

**TABLE A-5. Magnitude of SubScale Difference Required for .05 and .01 Level of Significance: Other**

| | Inf | | Com | | Ari | | Sim | | Voc | | Ds | | Pc | | Sp | | Pa | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| alpha | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 |
| Com | 2.6 | 3.5 | | | | | | | | | | | | | | | | |
| Ari | 3.2 | 4.1 | 2.7 | 3.6 | | | | | | | | | | | | | | |
| Sim | 2.8 | 3.6 | 2.2 | 2.9 | 2.8 | 3.7 | | | | | | | | | | | | |
| Voc | 3.3 | 4.3 | 2.9 | 3.8 | 3.4 | 4.4 | 3.0 | 3.9 | | | | | | | | | | |
| Ds | 3.2 | 4.2 | 2.8 | 3.7 | 3.3 | 4.3 | 2.9 | 3.8 | 3.4 | 4.5 | | | | | | | | |
| Pc | 3.3 | 4.2 | 2.8 | 3.7 | 3.3 | 4.3 | 2.9 | 3.8 | 3.4 | 4.5 | 3.4 | 4.4 | | | | | | |
| Sp | 3.0 | 4.0 | 2.6 | 3.4 | 3.1 | 4.1 | 2.7 | 3.5 | 3.2 | 4.2 | 3.2 | 4.2 | 3.2 | 4.2 | | | | |
| Pa | 3.5 | 4.6 | 3.1 | 4.1 | 3.6 | 4.7 | 3.2 | 4.2 | 3.7 | 4.9 | 3.7 | 4.8 | 3.7 | 4.8 | 0.0 | 0.0 | | |
| Oa | 3.0 | 3.9 | 2.5 | 3.2 | 3.0 | 4.0 | 2.6 | 3.4 | 3.2 | 4.1 | 3.1 | 4.0 | 3.1 | 4.1 | 0.0 | 0.0 | 3.4 | 4.5 |

Inf=Information; Com=Comprehension; Ari=Arithmetic; Sim=Similarities; Voc=Vocabulary; DS=Digit Symbol; PC=Picture Comprehension; SP=Spatial PA=Picture Arrangement; OA=Object Assembly

**TABLE A-6. Magnitude of SubScale Difference Required for .05 and .01 Level of Significance: Females**

| | Inf | | Com | | Ari | | Sim | | Voc | | Ds | | Pc | | Sp | | Pa | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| alpha | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 |
| Com | 2.4 | 3.1 | | | | | | | | | | | | | | | | |
| Ari | 2.8 | 3.7 | 2.4 | 3.1 | | | | | | | | | | | | | | |
| Sim | 2.4 | 3.1 | 1.8 | 2.4 | 2.4 | 3.1 | | | | | | | | | | | | |
| Voc | 3.1 | 4.0 | 2.7 | 3.5 | 3.1 | 4.0 | 2.7 | 3.5 | | | | | | | | | | |
| Ds | 2.9 | 3.8 | 2.5 | 3.3 | 2.9 | 3.8 | 2.5 | 3.3 | 3.2 | 4.2 | | | | | | | | |
| Pc | 2.8 | 3.7 | 2.4 | 3.1 | 2.8 | 3.7 | 2.4 | 3.1 | 3.1 | 4.0 | 2.9 | 3.8 | | | | | | |
| Sp | 3.1 | 4.1 | 2.8 | 3.6 | 3.1 | 4.1 | 2.8 | 3.6 | 3.4 | 4.4 | 3.2 | 4.2 | 3.1 | 4.1 | | | | |
| Pa | 3.1 | 4.1 | 2.7 | 3.6 | 3.1 | 4.1 | 2.7 | 3.6 | 3.4 | 4.4 | 3.2 | 4.2 | 3.1 | 4.1 | 3.4 | 4.5 | | |
| Oa | 2.8 | 3.7 | 2.4 | 3.1 | 2.8 | 3.7 | 2.4 | 3.1 | 3.1 | 4.0 | 2.9 | 3.8 | 2.8 | 3.7 | 2.8 | 3.7 | 3.1 | 4.1 |

Inf=Information; Com=Comprehension; Ari=Arithmetic; Sim=Similarities; Voc=Vocabulary; DS=Digit Symbol; PC=Picture Comprehension; SP=Spatial PA=Picture Arrangement; OA=Object Assembly

**TABLE A-7. Magnitude of SubScale Difference Required for .05 and .01 Level of Significance: Males**

| | Inf | | Com | | Ari | | Sim | | Voc | | Ds | | Pc | | Sp | | Pa | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| alpha | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 | 0.05 | 0.01 |
| Com | 2.5 | 3.2 | | | | | | | | | | | | | | | | |
| Ari | 3.0 | 3.9 | 2.5 | 3.3 | | | | | | | | | | | | | | |
| Sim | 2.6 | 3.3 | 2.0 | 2.7 | 2.6 | 3.4 | | | | | | | | | | | | |
| Voc | 3.1 | 4.0 | 2.6 | 3.5 | 3.1 | 4.1 | 2.7 | 3.6 | | | | | | | | | | |
| Ds | 3.0 | 4.0 | 2.6 | 3.4 | 3.1 | 4.0 | 2.7 | 3.6 | 3.2 | 4.2 | | | | | | | | |
| Pc | 2.9 | 3.8 | 2.5 | 3.3 | 3.0 | 3.9 | 2.6 | 3.4 | 3.1 | 4.0 | 3.1 | 4.0 | | | | | | |
| Sp | 3.0 | 4.0 | 2.6 | 3.4 | 3.1 | 4.0 | 2.7 | 3.5 | 3.2 | 4.2 | 3.2 | 4.1 | 3.1 | 4.0 | | | | |
| Pa | 3.1 | 4.1 | 2.7 | 3.5 | 3.2 | 4.1 | 2.8 | 3.6 | 3.3 | 4.3 | 3.2 | 4.2 | 3.1 | 4.1 | 3.2 | 4.2 | | |
| Oa | 2.8 | 3.6 | 2.3 | 3.0 | 2.8 | 3.7 | 2.4 | 3.2 | 2.9 | 3.8 | 2.9 | 3.8 | 2.8 | 3.7 | 2.9 | 3.8 | 3.0 | 3.9 |

Inf=Information; Com=Comprehension; Ari=Arithmetic; Sim=Similarities; Voc=Vocabulary; DS=Digit Symbol; PC=Picture Comprehension; SP=Spatial PA=Picture Arrangement; OA=Object Assembly